



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-482232

Comparative genomics of citric-acid producing *Aspergillus niger* ATCC 1015 versus enzyme-producing CBS 513.88 [working title: *Aspergillus niger* strain evolution]

S. Lucas, I. Grigoriev, H. Shapiro, A. Aerts, A. Atkins, A. Salamov, E. Lindquist, J. Pangilinan, Y. Lou, J. Grimwood, M. Andersen, M. Salazar, P. Schaap, P. van de Vondervoot, D. Culley, S. Baker, J. Thykaer, J. Frisvad, K. Nielsen, R. Albang, K. Albermann, R. Berka, G. Braus, S. Braus-Stromeyer, L. Corrochano, Z. Dai, P. van Dijck, G. Hofmann, L. Lasure, J. Magnusson, S. Meijer, J. Nielsen, M. Nielsen, A. van Ooyen, K. Panther, H. Pel, L. Poulsen, R. Samson, H. Stam, A. Tsang, J. van den Brink, C. Kubicek, D. Martinez, N. van Peij, J. Roubos, J. Nielsen

May 2, 2011

Genome Research

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

*Comparative genomics of citric-acid producing *Aspergillus niger* ATCC 1015 versus enzyme-producing CBS 513.88*

Mikael R. Andersen^a, Margarita P. Salazar^{a,b}, Peter J. Schaap^c, Peter J.I. van de Vondervoort^d, David Culley^e, Jette Thykaer^a, Jens C. Frisvad^a, Kristian F. Nielsen^a, Richard Albang^f, Kaj Albermann^f, Randy M. Berka^g, Gerhard H. Braus^h, Susanna A. Braus-Stromeier^h, Luis M. Corrochanoⁱ, Ziyu Dai^e, Piet W.M. van Dijck^d, Gerald Hofmann^j, Linda L. Lasure^e, Jon K. Magnusson^e, Susan L. Meijer^a, Jakob B. Nielsen^a, Michael L. Nielsen^a, Albert J.J. van Ooyen^d, Kathryn S. Panther^e, Herman J. Pel^d, Lars Poulsen^a, Rob A. Samson^k, Hein Stam^d, Adrian Tsang^l, Johannes M. van den Brink^m, Alex Atkinsⁿ, Andrea Aertsⁿ, Harris Shapiroⁿ, Jasmyn Pangilinanⁿ, Asaf Salamovⁿ, Yigong Louⁿ, Erika Lindquistⁿ, Susan Lucasⁿ, Jane Grimwood^o, Igor V. Grigorievⁿ, Christian P. Kubicek^p, Diego Martinez^{q,r}, Noël N.M.E. van Peij^d, Johannes A. Roubos^d, Jens Nielsen^{a,b}, and Scott E. Baker^{*,e}

^aCenter for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark.

^bCurrent address: Systems Biology, Department of Chemical and Biological Engineering, Chalmers University of Technology, SE-41296 Göteborg, Sweden.

^cSection Fungal Genomics, Laboratory of Microbiology, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands.

^dDSM Biotechnology Center, PO Box 1, 2600 MA Delft, The Netherlands.

^eFungal Biotechnology Team, Pacific Northwest National Laboratory, 99352 Richland, Washington, USA.

^fBiomax Informatics AG, Lochhamer Str. 9, D-82152 Martinsried, Germany.

^gNovozymes Inc, 1445 Drew Avenue, Davis, California 95616, USA.

^hMolecular Microbiology & Genetics, Georg-August University, Grisebachstrasse 8, D-37077 Goettingen, Germany.

ⁱDepartamento de Genetica, Universidad de Sevilla, Av. Reina Mercedes 6, E-41080 Sevilla, Spain.

^jNovozymes A/S, Hallas Alle 1, BD3.44, DK-4400 Kalundborg, Denmark.

^kCBS Fungal Biodiversity Centre, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.

^lCentre for Structural and Functional Genomics, Concordia University, 7141 Sherbrooke Street West, Montreal, QC, Canada H4B 1R6.

^mChr. Hansen A/S, Bøge Allé 10-12, DK-2970 Hørsholm, Denmark.

ⁿUS Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598, USA.

^oJoint Genome Institute, Stanford Human Genome Center, 975 California Avenue, Palo Alto, California 94304, USA.

^pResearch Area of Gene Technology and Applied Biochemistry, Institute of Chemical Engineering, Vienna University of Technology, Getreidemarkt 9-1665, A-1060 Vienna, Austria.

^qLos Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

^rDepartment of Biology, University of New Mexico, Albuquerque, New Mexico 87131, USA.

**To whom correspondence may be addressed. E-mail: Scott E. Baker, scott.baker@pnl.gov.*

28 April 2011

ACKNOWLEDGMENTS:

This study was partially funded by the Danish Research Agency for Technology and Production (MRA, SLM, JBN, MLN) and the National Council of Research Conacyt-Mexico (MPS). Genome sequencing was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396. Preparation of genomic DNA and some analysis of the genome were supported by the DOE Office of the Biomass Program.

DISCLAIMER:

[LLNL] This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus,

product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes

[LBNL] This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California

1 **Comparative genomics of citric-acid producing *Aspergillus***
2 ***niger* ATCC 1015 versus enzyme-producing CBS 513.88**

3 Mikael R. Andersen^a, Margarita P. Salazar^{a,b}, Peter J. Schaap^c, Peter J.I. van de Vondervoort^d,
4 David Culley^e, Jette Thykaer^a, Jens C. Frisvad^a, Kristian F. Nielsen^a, Richard Albang^f, Kaj
5 Albermann^f, Randy M. Berka^g, Gerhard H. Braus^h, Susanna A. Braus-Stromeier^h, Luis M.
6 Corrochanoⁱ, Ziyu Dai^e, Piet W.M. van Dijck^d, Gerald Hofmann^j, Linda L. Lasure^e, Jon K.
7 Magnusson^e, Susan L. Meijer^a, Jakob B. Nielsen^a, Michael L. Nielsen^a, Albert J.J. van Ooyen^d,
8 Kathryn S. Panther^e, Herman J. Pel^d, Lars Poulsen^a, Rob A. Samson^k, Hein Stam^d, Adrian
9 Tsang^l, Johannes M. van den Brink^m, Alex Atkinsⁿ, Andrea Aertsⁿ, Harris Shapiroⁿ, Jasmyn
10 Pangilinanⁿ, Asaf Salamovⁿ, Yigong Louⁿ, Erika Lindquistⁿ, Susan Lucasⁿ, Jane Grimwood^o,
11 Igor V. Grigorievⁿ, Christian P. Kubicek^p, Diego Martinez^{q,r}, Noël N.M.E. van Peij^d, Johannes
12 A. Roubos^d, Jens Nielsen^{a,b}, and Scott E. Baker^{*,e}.

13

14 ^aCenter for Microbial Biotechnology, Department of Systems Biology, Technical University of
15 Denmark, DK-2800 Kgs. Lyngby, Denmark. ^bCurrent address: Systems Biology, Department of
16 Chemical and Biological Engineering, Chalmers University of Technology, SE-41296 Göteborg,
17 Sweden. ^cSection Fungal Genomics, Laboratory of Microbiology, Wageningen University,
18 Dreijenplein 10, 6703 HB Wageningen, The Netherlands. ^dDSM Biotechnology Center, PO Box 1, 2600
19 MA Delft, The Netherlands. ^eFungal Biotechnology Team, Pacific Northwest National Laboratory,
20 99352 Richland, Washington, USA. ^fBiomax Informatics AG, Lochhamer Str. 9, D-82152 Martinsried,
21 Germany. ^gNovozymes Inc, 1445 Drew Avenue, Davis, California 95616, USA. ^hMolecular
22 Microbiology & Genetics, Georg-August University, Grisebachstrasse 8, D-37077 Goettingen,
23 Germany. ⁱDepartamento de Genetica, Universidad de Sevilla, Av. Reina Mercedes 6, E-41080 Sevilla,

1 Spain. ^jNovozymes A/S, Hallas Alle 1, BD3.44, DK-4400 Kalundborg, Denmark. ^kCBS Fungal
2 Biodiversity Centre, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands. ^lCentre for Structural and
3 Functional Genomics, Concordia University, 7141 Sherbrooke Street West, Montreal, QC, Canada
4 H4B 1R6. ^mChr. Hansen A/S, Bøge Allé 10-12, DK-2970 Hørsholm, Denmark. ⁿUS Department of
5 Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598, USA. ^oJoint
6 Genome Institute, Stanford Human Genome Center, 975 California Avenue, Palo Alto, California
7 94304, USA. ^pResearch Area of Gene Technology and Applied Biochemistry, Institute of Chemical
8 Engineering, Vienna University of Technology, Getreidemarkt 9-1665, A-1060 Vienna, Austria. ^qLos
9 Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. ^rDepartment of Biology,
10 University of New Mexico, Albuquerque, New Mexico 87131, USA.

11

12 *Corresponding author:

13 Scott E Baker,

14 Pacific Northwest National Lab,

15 902 Batelle Blvd,

16 Richland WA 99352

17 Phone: (509) 372-4759

18 Fax: (509) 372-4732

19 e-mail: scott.baker@pnl.gov

20

21 Running title: *Aspergillus niger* strain evolution

22

23 Keywords: Biotechnology, *Aspergillus*, comparative genomics, physiology, transcriptomics.

24

1 **Abstract**

2 The filamentous fungus *Aspergillus niger* exhibits great diversity in its phenotype. It is found
3 globally, both as marine and terrestrial strains, produces both organic acids and hydrolytic
4 enzymes in high amounts, and some isolates exhibit pathogenicity. Although the genome of
5 an industrial enzyme-producing *A. niger* strain (CBS 513.88) has already been sequenced,
6 the versatility and diversity of this species compels additional exploration. We therefore
7 undertook whole genome sequencing of the acidogenic *A. niger* wild type strain (ATCC
8 1015), and produced a genome sequence of very high quality. Only 15 gaps are present in
9 the sequence and half the telomeric regions have been elucidated. Moreover, sequence
10 information from ATCC 1015 was utilized to improve the genome sequence of CBS 513.88.
11 Chromosome-level comparisons uncovered several genome rearrangements, deletions, a
12 clear case of strain-specific horizontal gene transfer, and identification of 0.8 megabase of
13 novel sequence. Single nucleotide polymorphisms per kilobase (SNPs/kb) between the two
14 strains were found to be exceptionally high (average: 7.8, maximum: 160 SNPs/kb). High
15 variation within the species was confirmed with exo-metabolite profiling and phylogenetics.
16 Detailed lists of alleles were generated, and genotypic differences were observed to
17 accumulate in metabolic pathways essential to acid production and protein synthesis. A
18 transcriptome analysis revealed up-regulation of the electron transport chain, specifically
19 the alternative oxidative pathway in ATCC 1015, while CBS 513.88 showed significant up-
20 regulation of genes relevant to glucoamylase A production, such as tRNA-synthases and
21 protein transporters.

1 Our results and datasets from this integrative systems biology analysis resulted in a
2 snapshot of fungal evolution and will support further optimization of cell factories based on
3 filamentous fungi.

4 [Supplemental materials (10 figures, three text documents and 16 tables) have been made
5 available. The whole genome sequence for *A. niger* ATCC 1015 is available from NCBI under
6 acc. no ACJ000000000. The up-dated sequence for *A. niger* CBS 513.88 is available from
7 EMBL under acc. no AM269948-AM270415. The sequence data from the phylogeny study
8 has been submitted to NCBI (GU296686-296739). Microarray data from this study is
9 submitted to GEO as series GSE10983. Accession for reviewers is possible through:
10 <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=dxurfocwimogqby&acc=GSE10983>
11 The dsmM_ANIGERa_coll511030F library and platform information is deposited at GEO
12 under number GPL6758.

13

1 **Introduction**

2 The saprotrophic filamentous fungus *Aspergillus niger* is found globally, both as marine and
3 terrestrial strains and exhibits a great diversity in its phenotype. *A. niger* has become one of
4 the major workhorses in industrial biotechnology, being very efficient in producing both
5 polysaccharide-degrading enzymes (particularly amylases, pectinases and xylanases) or
6 organic acids (mainly citric acid) in high amounts. It also has a long history of safe use
7 (Schuster et al. 2002; van Dijck et al. 2003; van Dijck 2008). Commercial importance is
8 illustrated by a world market for industrial enzymes of nearly US\$ 5 billion in 2009, of which
9 filamentous fungi account for roughly half of the production (Lubertozzi and Keasling 2008),
10 and a global citric acid production of 9×10^6 metric tons in 2000 (Karaffa and Kubicek 2003).
11 In 2007, the genome sequence of *A. niger* strain CBS 513.88, used for industrial enzyme-
12 production, was published (Pel et al. 2007). This strain was derived from *A. niger* NRRL 3122,
13 a strain developed for glucoamylase A production by classical mutagenesis and screening
14 methods (van Lanen and Smith 1968). This work initiated a number of new genome-based
15 investigations (Sun et al. 2007; Yuan et al. 2008ab; Andersen et al. 2008ab; Martens-
16 Uzunova and Schaap 2008), but did not reveal differences between citric-acid-producing
17 and enzyme-producing *A. niger* strains (Cullen 2007).
18 In this paper, we present the nearly complete genome sequence of the citric acid-producing
19 *A. niger* wild type strain ATCC 1015 and compare it to the genome sequence of the enzyme-
20 producing strain CBS 513.88. The genetic diversity of these two *A. niger* strains was
21 determined by applying systems biology tools as well as new bioinformatic methods to

1 examine multi-level differences that distinguish the wild-type citric acid-producing strain
2 from the mutagenized glucoamylase A-producing strain.

3 ***Results***

4 **General genome statistics**

5 The 34.85 Mb genome sequence of *A. niger* ATCC 1015 was generated using a shotgun
6 approach and then further improved to a high-quality assembly of 24 finished contigs
7 separated by 15 gaps (including 8 from centromeric regions). Genome statistics are
8 summarized in Table 1, with details in Suppl. Text 1. The full sequence and annotations are
9 available from the Joint Genome Institute (JGI) Genome Portal ([http://genome.jgi-
10 psf.org/Aspni5](http://genome.jgi-psf.org/Aspni5)) and from NCBI (Acc. no. ACJE000000000).

11 The genome sequence of *A. niger* CBS 513.88 (Pel et al., 2007) was improved using the ATCC
12 1015 sequence to close 186 contig gaps and modify the gene models associated with these
13 gaps (Table 1 and Suppl. Table 1). The updated *A. niger* CBS 513.88 genome sequence is
14 accessible through EMBL (Acc.nos AM269948-AM270415).

15 We note a large difference (2,882) in the number of called genes in the two strains (Table 1).
16 A thorough analysis indicates an over-prediction of genes in CBS 513.88 and an under-
17 prediction of genes in ATCC 1015, and 396/510 unique genes in CBS 513.88/ATCC 1015
18 (Suppl. Text 2, Details in Suppl. Tables 2-9, and Suppl. Figure 1).

1 Unique genes in both strains suggest horizontal gene transfer to be a cause
2 for the amylase hyper-producer phenotype.

3 For the genes unique to the amylase producing strain CBS 513.88 (Suppl. Table 6), the most
4 notable genes are two α -amylases which are identical to the *Aspergillus oryzae* α -amylase, a
5 possible cause for the amylase hyper-producer phenotype. We discuss this in detail in a
6 later section (Figure 2). Furthermore, we find three possible polyketide synthases, which
7 suggests a unique secondary metabolite profile for this strain.

8 Examining the genes found only in the ATCC 1015 (Suppl. Table 6), does not point to any
9 obvious cause for citric acid hyper-production, but four possible polyketide synthases and a
10 putative NRPS are found to be unique for ATCC 1015, suggesting this strain to have unique
11 secondary metabolites as well. Multiple effects of this type are indeed seen for both strains
12 in analyses later in the article (Figure 1, Table 2, Suppl. Figure 10).

13 Synteny mapping shows 0.5 Mb of chromosomal rearrangements and a
14 whole-arm inversion

15 The genomes of the two strains are largely syntenic (Figure 1). For example, the 1,429
16 protein-encoding genes of CBS 513.88 supercontig An01 that have predicted counterparts in
17 ATCC 1015 are, with one exception, all mapped to chromosome 2b. A similar example is
18 seen for all but one of 1,486 protein-encoding genes on supercontig An02. Both exceptions
19 encode putative Tan1-like transposases. Notwithstanding, a number of significant
20 differences in genome configuration exist (Figure 1). More than 0.5 Mb of additional
21 genome sequence in ATCC 1015 resides in four large regions on chromosomes (Chrs) II, III
22 and V (Figure 1). The flanking sequences of three of these additional elements in ATCC 1015

1 are syntenic to continuous sequences in CBS 513.88; and are thus true differences in
2 genome configuration (Suppl. Table 2 and Suppl. Figure 2, details on Chr III in the next
3 section). The fourth region on the left arm of Chr V in ATCC 1015 could not be verified due
4 to a contig gap for CBS 513.88. The contig gap in the right arm of Chr V in the genome
5 sequence of ATCC 1015 is spanned by continuous sequence in CBS 513.88, containing 15
6 predicted genes. An overview of the genes found in the gaps unique to ATCC 1015 can be
7 found in Suppl. Table 2.

8 Other striking differences include a large inversion in Chr VIII and the inversion and
9 translocation of a large fragment between the left arms of Chrs III and VII (Suppl. Figure 3).
10 Both were confirmed with PCR spanning the breakpoints (data not shown).
11 Finally, the presence of telomere sequences in genome data confirms an inversion of the
12 complete right arm of Chr VI.

13 **Two extra α -amylase encoding genes likely obtained through HGT**

14 An un-matched region identified for the left arm of Chr III (Figure 1) spans 72.5 kb of unique
15 sequence in ATCC 1015. Remarkably, for CBS 513.88 a unique 85.3 kb sequence is found at
16 this location (Figure 2, Panel A). Gene annotation may be found in Suppl. Table 10. The
17 absence of the region in the chromosome of ATCC 1015 was confirmed experimentally by
18 gDNA hybridizations (Suppl. Figure 4, Suppl. Table 16).

19 Approximately 12 kb of this region shares more than 99.8 % sequence identity at DNA level
20 with genomic DNA from *Aspergillus oryzae* RIB40 (Figure 2, panel B). The complete 85 kb
21 fragment including the 12 kb α -amylase region also appears to be involved in a recent
22 duplication recombination event to Chr VII, (Figure 2, panel C). Thus, the CBS 513.88

1 genome harbors two additional α -amylase encoding genes which are orthologues to the α -
2 amylase encoding genes (AO090023000944, AO090120000196) of *A. oryzae* RIB40. These
3 findings suggest that CBS 513.88 and the parental strain NRRL 3122 (data not shown)
4 acquired these duplicate α -amylase genes (An12g06930, An05g02100) through horizontal
5 gene transfer (HGT). The occurrence of HGT in *Aspergilli* would further be supported by the
6 presence of α -amylase-encoding genes in other black *Aspergilli* which display over 99 % nt
7 identity to those of *A. oryzae* RIB40 and the *A. niger* CBS 513.88 α -amylases (Korman et al.
8 1990; Shibuya et al. 1992). The origin of the remaining part of the un-matched region
9 remains unclear. No significant similarity was observed at DNA level and only a few of the
10 encoded proteins show similarity with other proteins present in the NR protein database.
11 There is also evidence that this HGT may have occurred by the action of transposases. The
12 12 kb HGT region is flanked by 202-bp inverted, perfect, terminal repeats (ITR) (Figure 2,
13 Panels A, B). Furthermore, this HGT region harbors another gene, An12g07000, which is
14 identical to the *A. oryzae tnpA* transposase gene. Interestingly, in the *A. oryzae* RIB40
15 genome, the same 12 kb fragment has been duplicated twice and is present on multiple
16 chromosomes, but only one perfect copy of the ITRs is retained (Figure 2, Panel C).

17 **Transposon presence is strain-specific**

18 Transposon-like sequences were identified in both genomes and quantified (Suppl. Table
19 11). This comparison pinpoints a remarkable difference in presence and amount of
20 transposon-related sequences in ATCC 1015 and CBS 513.88 both for class I and for class II
21 transposons: only 16 sequences were identified in ATCC 1015 but 55 were detected in CBS
22 513.88. Whereas the ATCC 1015 genome contains no class I superfamily copia and a single

1 copy of class II superfamily Fot1/pogo, these sequences are much more abundant in CBS
2 513.88, where 15 class I copia, and 13 class II Fot1/pogo are found.

3 **SNP analysis reveals high mutation rates and hypervariable regions**

4 We found 8 ± 16 SNPs/kb (average \pm standard deviation) and a maximum of 163 SNPs/kb
5 single nucleotide polymorphisms (SNPs) between *A. niger* strains ATCC 1015 and CBS 513.88
6 (Suppl. Table 12). This value is much higher than the maximum of 9 SNPs/kb found by
7 Cuomo et al. (2007) in a SNP-analysis of two *Fusarium graminearum* strains. A comparison
8 of the *A. niger* ATCC 1015 genome sequence to that of *A. niger* ATCC 9029 revealed
9 markedly less variation between the two (2 ± 5 SNPs/kb) (Suppl. Figure 5), indicating a large
10 genomic variation in the *A. niger* group, but little between ATCC 1015 and 9029. These
11 polymorphisms are not uniformly distributed, but cluster in hyper-variable regions (Figure
12 1). This is supported by a gDNA hybridization study, which confirms the presence of distinct
13 hyper-variable regions (Suppl. Figure 4).

14 **Gene comparisons reveal industrially relevant strain-differences**

15 To identify strain-defining systemic effects of the genome variation, results from the Imprint
16 gene synteny analysis (See methods and Suppl. Table 5) were used for additional genome-
17 scale investigation.

18 GO term over-representation analysis was performed on gene groups with distinct common
19 properties (Group overview in Suppl. Table 5, details in Suppl. Table 13). Most interesting is
20 the observation that 37 proteins involved in transcriptional regulation are non-functional in
21 CBS 513.88 due to frameshifts or stop codons. This suggests a less stringent regulation of
22 CBS 513.88 relative to ATCC 1015. A comparative shake-flask study of the two strains also

1 suggests mutations in regulatory elements as nitrogen source utilization is impaired in the
2 CBS 513.88 strain (data not shown). While the nitrogen catabolism regulator AreA was
3 originally thought to be mutated in CBS 513.88 (Pel et al. 2007), a re-sequencing of *areA*
4 showed the ORFs to be identical.

5 To detect potential differences in the metabolism of the two strains, we compared all genes
6 encoding proteins that display differences at the amino acid level between the two strains
7 on the metabolic network of *A. niger* (Andersen et al. (2008a) and mapped them to
8 metabolic pathways (Suppl. Figure 6). Mutations were found in the pathways for
9 biosynthesis of proline, aspartate, asparagine, tryptophan and histidine, which may be
10 relevant to protein production. Also, mutations were found in the plasma membrane-bound
11 ATPase, in the enzymes of the GABA shunt, of the TCA cycle and in components of all steps
12 of the electron transport chain, which could be relevant for the production of citric acid.

13 **Phylogenetic analysis confirms high variability between genome sequences**

14 To place the comparison of *A. niger* ATCC 1015 and CBS 513.88 into the larger context of the
15 *Aspergillus* section *Nigri*, variation in the two strains was analyzed by phylogenetic analysis
16 of a part of the β -tubulin sequence for a number of *A. niger* strains and other black *Aspergilli*
17 (Figure 3A). To further explore the genome variability across the *A. niger* species, we
18 identified four 1-kb regions from Chrs II, IV, VI, and VIII that were identical in the ATCC 1015
19 and ATCC 9029 strains, but had approximately 20 SNPs/kb relative to the genome sequence
20 of CBS 513.88. These regions were PCR-sequenced in seven *A. niger* strains including the
21 CBS 513.88 progenitor, NRRL 3122 and the *A. niger* neotype strain CBS 554.65 (Kozakiewicz
22 et al. 1992) to form a phylogenetic tree with high variation (Figure 3B). For previously
23 sequenced strains, the re-sequencing results were identical to the previously determined

1 genomic sequence, thereby excluding that the genomic variation is an artifact of low-fidelity
2 sequencing.

3 All *A. niger* strains have an identical sequence for β -tubulin and form a strongly supported
4 terminal clade together with *A. awamori* (Figure 3A). The other taxonomically different
5 species are phylogenetically distinct from this group which is in accordance with the serial
6 classification of Frisvad et al. (2007a). The latter clustering based on the selected regions
7 (Figure 3B) was able to separate the *A. niger* strains into three groups. Strain CBS 513.88 and
8 its progenitor NRRL 3122 were identical for the regions, indicating limited SNP introduction
9 due to classical strain improvement, and thus confirms high variation in the *A. niger* group.

10 **Exo-metabolomic profiling describes three distinct groups of *A. niger***

11 To further compare the two sequenced strains on a system-wide, but non-genomic level, to
12 each other and to other members of the *A. niger* species group, exo-metabolomic profiles of
13 the two strains were compared to nine other strains, including the CBS 513.88 progenitor
14 strain NRRL 3122, the widely used laboratory strain ATCC 9029, and the *A. niger* neotype
15 CBS 554.65 (Table 2). This gave three distinct clades, which follow the clustering of Figure
16 3B, but does not conform to the phylogenetic distances calculated. Strain CBS 513.88 and its
17 progenitor strain had very similar secondary metabolites profiles. These two strains differed
18 from the other strains analyzed. Strain ATCC 1015 had a profile similar to seven other strains
19 although there were some quantitative differences. Strain CBS 126.49 had a unique profile.

20 In the case of ochratoxin A, it is interesting that in ATCC 1015 and ATCC 9029, a remnant of
21 the PKS gene (An15g07920) of the putative ochratoxin cluster in *A. niger* CBS 513.88 (Pel et
22 al. 2007) was identified, which could be related to a 21-kb deletion in both strains (Suppl.
23 Figure 7). We have noted that a putative NRPS is found in one of the unique chromosome

1 regions found in ATCC 1015 (Figure 1), which may account for some of the difference (Suppl.
2 Table 2).

3 **Transcriptome analysis of *A. niger* ATCC 1015 and CBS 513.88 growing on** 4 **glucose**

5 To evaluate the effect of the differences in genome sequence on the physiology of the two
6 strains, batch-cultivations in bioreactors and a comparative transcriptome analysis was
7 performed. Strains ATCC 1015 and CBS 513.88 were grown under the same conditions in
8 batch cultures in a glucose-based minimal medium. The GlaA-producing strain CBS 513.88
9 produced more than 1.2 g/L GlaA more than ATCC 1015, while producing approximately 1
10 g/L biomass less. Other measured characteristics were similar for the two strains (Table 3).
11 Statistical analysis showed 4,784 significantly ($\text{adj.p} < 0.05$) differentially expressed genes,
12 with an almost equal number of genes up-regulated in either strain (2,431 in CBS 513.88 vs
13 2,353 for ATCC 1015).
14 Examining the differential expression in the context of metabolic pathways (Suppl. Figure 8),
15 only the alternative oxidative pathway has uniformly higher expression in ATCC 1015, while
16 a substantial subset of metabolism was up-regulated in CBS 513.88, including glycolysis and
17 the TCA cycle. As the specific growth rates of the strains are similar, we suggest that this
18 extra activity of central metabolism provides precursors for the higher productivity of
19 glucoamylase. We also find increased expression of genes involved in amino acid
20 metabolism, especially the entire biosynthetic pathways of threonine, serine, and
21 tryptophan. Intriguingly, analysis of the amino acid composition of the glucoamylase A
22 protein (Suppl. Figure 9) revealed that GlaA is atypical in that it has a higher content of

1 specifically tryptophan, threonine, and serine than 90% of the predicted genes of CBS
2 513.88. For all three amino acids, the content is almost twice as high as the average in the
3 composition of total protein (Christias et al., 1975).

4 We initially thought that this was due to an identified frame shift mutation identified in the
5 gene for the general amino acid regulating transcription factor (cross-pathway control
6 protein) CpcA (Suppl. Table 5), but re-sequencing of this gene showed this frame shift to be
7 a sequencing error. The presence of two sense mutations was confirmed and is currently
8 being investigated.

9 To further support that the increased production of GlaA is significant enough to affect
10 amino acid biosynthesis, we used a genome-scale metabolic model of *A. niger* (Andersen et
11 al. 2008a) to model the two strains under the growth conditions used (Table 3). The
12 computed fluxes (Suppl. Table 14) show that for all three amino acids, the fluxes through
13 the biosynthetic pathways must be at least twice as high in CBS 513.88 compared to ATCC
14 1015 to support the increased GlaA production, thus corresponding well with the
15 transcriptome results.

16 For a broader analysis of trends revealed by the transcriptome profiles, a GO term
17 overrepresentation analysis was conducted on the genes significantly up-regulated in either
18 strain (Suppl. Table 15). The analysis confirmed that traits relevant for protein production –
19 such as specifically amino acid biosynthesis and tRNA aminoacylation - appear to be highly
20 significant for CBS 513.88. For ATCC 1015, GO biological functions for electron transport
21 (adj.p = 2.7e-05), carbohydrate transport (adj.p. = 0.0068), and organic acid transport (adj.p
22 = 0.035) were significant.

1 An examination of expression of individual genes showed that *glaA* had significantly
2 increased expression in CBS 513.88 (adj.p = 84e-6), but the fold change (3.2) was lower than
3 the increase in enzyme activity (6-fold, Table 3). Interestingly, all identified tRNA-aminoacyl
4 synthases were found to be two- to six-fold up-regulated in CBS 513.88.

5 **Mapping of gene expression to the genome identifies secondary metabolite** 6 **cluster activities and underpins the whole-arm inversion in chromosome VI**

7 To examine differences in gene expression between the two strains relative to chromosome
8 positions, the log2-ratios of the gene expression indices from the transcriptome analysis
9 were mapped to the synteny diagrams of Figure 1 (Suppl. Figure 10).

10 The analysis of this mapping identified a number of features. First, six regions not found in
11 CBS 513.88, contain genes with uniformly higher expression in ATCC 1015, supporting the
12 absence of these regions in CBS 513.88. Second, two active secondary metabolite clusters
13 were identified on Chr VIII (contains a polyketide synthase (ID: 211885) which is unique for
14 ATCC 1015) and Chr I (NRPS cluster, found in both strains (NRPS ID: 43555/An09g00520),
15 but appears only to be active in CBS 513.88). Third, the inversion of the entire right arm of
16 Chr VI (Figure 1) is further supported. The telomeric position effect, which defines
17 decreased expression in the vicinity of telomeres, has been described for *S. cerevisiae*
18 (Gottschling et al. 1990). Thus, if an arm has been inverted, reduced expression should be
19 found at opposite ends in ATCC 1015 and CBS 513.88. This is indeed reflected in the log2-
20 ratio on the right arm of Chr VI (Suppl. Figure 10).

1 **Discussion**

2 In this study, we provide and compare the genomes of two strains of *A. niger*. These two
3 strains have very different phenotypes, one, the predecessor to efficient enzyme-producing
4 strains having undergone some level of mutagenesis and selection, and the other a wild-
5 type and parent strain of high citric acid producing strains. This makes the comparison
6 interesting both in terms of basic science and industrial applications. We have supported
7 the conclusions of our comparison with further experiments, allowing us to propose new
8 hypotheses and conclusions within three main areas; genetic diversity of the *A. niger* group,
9 horizontal gene transfer in fungi, and fungal biotechnology (Discussed separately below).

10 The diversity of the two strains was explored through a multi-level (DNA, chromosome,
11 gene and protein) comparison between both genome architectures. Interestingly we see
12 both a remarkable similarity at the level of chromosome, gene order and gene identity, but
13 also notable diversity, specifically high levels of SNPs, a 0.8 Mb difference in genome size,
14 several large insertion/deletions of up to 200 kb, different transposon populations, a major
15 translocation/inversion, the inversion of an entire chromosomal arm, and a large set of
16 unique genes in both species (points 1-5 below).

17 1) 4-500 unique genes were identified for each strain, most of which are evenly distributed
18 over the chromosomes. This indicates strain evolution by a high frequency of loss and/or
19 uptake of gene fragments. Interestingly, the opposite is seen in two strains of the
20 pathogenic *A. fumigatus*, where 80 % of the strain-specific genes (143 and 218 respectively)
21 are found in a few, large isolate-specific genomic islands (Fedorova et al. 2008). Thus, this
22 *intra*-species increased frequency of transfer/loss of genetic elements so far seems to be

1 specific for the *A. niger* group, more genome sequences may prove it to be present in other
2 *Aspergilli*.

3 2) The whole-arm inversion on Chr VI is a remarkable event. Over the characterized
4 *Aspergilli* sequenced genomes, most described inversions were isolated as mutants, with a
5 breakpoint in a gene, leading to a phenotype. Pel et al. (2007) also reported high synteny
6 between the centromeric parts of the arms of *A. niger* CBS 513.88 and *A. nidulans* FGSC A4.
7 Even so, in this case, it is supported by the genome sequence, the presence of telomeric
8 sequences and a detectable telomeric positioning effect. From Figure 2 in Galagan et al.
9 2005, it can also be seen that centromeric inversions must have occurred in the genealogy
10 of the common ancestor of *A. nidulans*, *A. oryzae*, and *A. fumigatus*.

11 3) The large translocation/inversion event of Chrs III and VII is interesting as it explains the
12 discrepancies in chromosome size reported by Pel et al. (2007) between strain N400
13 (Verdoes et al. 1994) and calculated sizes of Chrs III and VII for CBS 513.88. This fact and that
14 the breakpoints are within two otherwise intact genes indicates that the event has
15 happened after the branching of strains N400 and NRRL 3122, possibly in the mutagenesis
16 of the predecessor of CBS 513.88.

17 4) The presence and functionality of transposons is qualitatively and quantitatively less
18 complex in the genome of ATCC 1015 than in CBS 513.88 (Suppl. Table 11). The different
19 distribution of transposons between different strains has been observed before (Braumann
20 et al. 2008) and has also been described to be inducible by mutagenesis and other stress-
21 induced conditions (Strand et al. 1985). However, the scale of differences makes a recent
22 mutagenesis program unlikely to be the basis for the multiple HGT events. Furthermore, the
23 detection of repeat sequences in a number of DNA regions was accompanied by strain-

1 specific differences. We therefore suggest that the transposon populations have played a
2 significant role in strain evolution.

3 5) Finally, the phylogenetic analysis (Figure 3B) confirms the high genetic variation described
4 for two strains to be general within the *A. niger* group, and not the result of either CBS
5 513.88 or ATCC 1015 being atypical. This is again supported by the exo-metabolomic profiles
6 of the 11 examined *A. niger* strains (Table 2). While the grouping of the exo-metabolites
7 does not accurately reflect the cladograms in Figure 3, we see this as confirmation of the
8 high genetic variation in the *A. niger* group, as the production of a given exo-metabolite may
9 be changed by a single genomic event. Thus, we believe that the diversity of the two
10 genome-sequenced strains is common for the *A. niger* group, which seems to have highly
11 dynamic genomes.

12 The presence/absence of HGT in fungi have often been discussed (Galagan et al. 2005;
13 Keeling and Palmer et al. 2008; Khaldi and Wolfe 2008; Khaldi et al. 2008), but in this study
14 we believe to have shown HGT to be the most likely origin of two α -amylase genes in CBS
15 513.88 that have seemingly been transferred to *A. niger* from another species, possibly *A.*
16 *oryzae*. The HGT must have occurred before the separation of CBS 513.88 and ATCC 1015.
17 The amylase genes are not found in the ATCC 1015 strain, but they are in CBS 513.88
18 flanked by a transposon, which is found both in *A. oryzae* RIB40 and in a truncated form in
19 ATCC 1015. With the evidence of HGT into *A. clavatus* from a *Magnaporthe*-related donor,
20 presented by Khaldi et al. (2008), HGT is now identified in two separate cases with distantly
21 related species, and may thus be seen as a general phenomenon in filamentous fungi.

22 A large part of the reason for comparing the two strains, has been to gain insight in the
23 genetic basis for the two industrially relevant phenotypes. In citric acid-producing ATCC

1 1015, we are intrigued to see higher expression of elements from the electron transport,
2 carbohydrate transport, and organic acid transport. These factors are known to be involved
3 in obtaining high citric acid yields. Higher expression levels even though citric acid
4 production is not notably higher (see discussion below), suggests that these traits may have
5 been contributing factors to the original selection of ATCC 1015 for citric acid production.
6 For the GlaA-producing strain (CBS 513.88), we observe systemic changes: mutations in
7 regulatory genes (e.g. *cpcA*), higher expression of *glaA* itself, and up-regulation of all
8 identified tRNA-synthases may all contribute to a more efficient enzyme producer. We also
9 present the hypothesis that increased production of the amino acids serine, threonine and
10 tryptophan may be required for efficient GlaA production, as these are overrepresented in
11 GlaA.

12 While we have identified multiple possible contributors to the efficient amylase-production
13 of CBS 513.88, we have only seen a few factors associated with citric acid production. We
14 propose multiple reasons for this: 1) CBS 513.88 was selected for increased GlaA production
15 while ATCC 1015 is a wild type strain with only modest citric acid production (compared to
16 current yields). 2) The medium used for the transcriptome analysis is not optimized for citric
17 acid production, as this would not give the dispersed growth necessary for representative
18 sampling (Karaffa et al. 2003). 3). Protein synthesis is a complex process, with more factors
19 involved than in the production and secretion of a simple organic acid. Since more than
20 6000 genes differ on the amino acid level between the two strains, it would be unlikely not
21 to find many differences involved in protein production. For this reason, we have put our
22 emphasis on genes found in multiple types of analysis, effects in entire pathways, and traits
23 (GO) that are found to be statistically overrepresented, which allows for robust conclusions.

1 In conclusion, our results establish a firm comparative genomics foundation on which to
2 build and test hypotheses regarding enzyme production, organic acid production and
3 diversity within a “species group”.

4 ***Methods***

5 **ATCC 1015 genome assembly**

6 The sequence reads were derived from four whole-genome shotgun (WGS) libraries: One
7 with an insert size of 2–3 kbp, two with an insert size of 6–8 kbp, and one with an insert size
8 of 35–40 kbp. The reads were screened for vector using cross_match and then trimmed for
9 vector sequence and quality (J Chapman, N Putnam, I Ho and D Rokhsar, unpublished).
10 Reads shorter than 100 bases after trimming were excluded. The final data set included:
11 28,551 of 2–3 kb reads, containing 21.5 Mb of sequence; 160,479 of 6–8 kb reads,
12 containing 123 Mb of sequence; and 38,651 of 35–40 kb reads, containing 23.8 Mb of
13 sequence.

14 The data were assembled using release 2.7 of Jazz, a WGS assembler developed at the JGI (J
15 Chapman, N Putnam, I Ho and D Rokhsar, unpublished; Aparicio et al. 2002). The genome
16 size and sequence depth were initially estimated to be 36 Mb and 8.0, respectively. After
17 removal of short (less than 1 kb) and redundant scaffolds (shorter than 5 kb with 80% or
18 more of the length matching a scaffold longer than 5 kb), the assembly included 43.7 Mb of
19 scaffold sequence with 8.2 Mb (18.9%) of gaps in 350 scaffolds, with half of the scaffold
20 sequence contained in the eight largest scaffolds of 1.81 Mb or longer. The sequence depth
21 derived from the assembly was 7.88 ± 0.05 . To estimate the completeness of the assembly, a
22 set of 50,001 ESTs was BLAT-aligned to the unassembled trimmed reads, as well as the

1 assembly itself. 43,323 ESTs (86.6%) were more than 80% covered by the unassembled data,
2 43,978 (88.0%) were more than 50% covered, and 44,206 (88.4%) were more than 20%
3 covered. By way of comparison, 48,798 ESTs (97.6%) showed hits to the assembly.
4 Orientation of the chromosome arms was based on eight available telomere sequences, the
5 supercontig orientation proposed by Pel et al. (2007) and research on linkage groups in *A.*
6 *niger* (Bos et al. 1989; Debets et al. 1989, 1990ab; Swart et al. 1992; Verdoes et al. 1994).

7 **ATCC 1015 genome finishing**

8 To perform genome improvement on *A. niger* ATCC 1015, initial read layouts from the
9 whole genome shotgun assembly were converted into the Phred/Phrap/Consed pipeline
10 (Gordon et al. 1998). Following manual inspection of the assembled sequences, finishing
11 was performed by resequencing plasmid subclones and by walking on plasmid subclones or
12 fosmids using custom primers. All finishing reactions were performed with 4:1 BigDye to
13 dGTP BigDye terminator chemistry (Applied Biosystems). Repeats in the sequence were
14 resolved by transposon-hopping 8kb plasmid clones. Fosmid clones were shotgun
15 sequenced and finished to fill large gaps, resolve large repeats or to resolve chromosome
16 duplications and extend into chromosome telomere regions.

17 The resulting 24 finished scaffolds were orientated where possible into chromosome
18 structures using comparisons to Pel et al. (2007) and telomere positions. The improved
19 genome consists of 34,853,277 base pairs of with an estimated error rate of less than 1
20 error in 100,000 bp.

1 **ATCC 1015 automatic annotation**

2 Gene models in the genome of *A. niger* were predicted using Fgenesh (Salamov and
3 Solovyev 2000), Fgenesh+ (Salamov and Solovyev 2000), and Genewise (Birney and Durbin
4 2000) integrated in the JGI annotation pipeline. Fgenesh was trained on a set of over 2,000
5 putative full-length transcripts derived from clustered *A. niger* ESTs and reliable homology-
6 based gene models to show 81% sensitivity and 81% specificity of predictions on a test set.
7 Homology based gene predictors were seeded with blastx alignments of proteins from the
8 NCBI non-redundant set of proteins. 31,578 *A. niger* ESTs were sequenced using Sanger
9 technology and were either directly mapped to genomic sequence when the ESTs included
10 putative full-length (FL) genes, or used to extend predicted gene models into FL genes by
11 adding 5' and/or 3' UTRs. In addition, 386,515 ESTs with an average length of 104
12 nucleotides were sequenced with 454 GS20 sequencers and used in validation of predicted
13 gene models. Since multiple gene models were generated for each locus, a single
14 representative model was chosen based on homology and EST support and used for further
15 analysis.

16 All predicted gene models were functionally annotated by sequence similarity to annotated
17 genes from the NCBI non-redundant set and other specialized databases (such as KEGG
18 (Kanehisa et al. 2002; Kanehisa et al. 2004)) using Blast and hardware accelerated double-
19 affine Smith-Waterman alignments (<http://www.timelogic.com>). Functional and structural
20 domains were predicted in protein sequences using the InterPro software (Zdobnov and
21 Apweiler 2001). All genes were also annotated according to Gene Ontology (Ashburner et al.
22 2000; Harris et al. 2004), eukaryotic orthologous groups (KOGs (Koonin et al. 2004)), and
23 KEGG metabolic pathways (Kanehisa et al. 2004).

1 **ATCC 1015 sequence availability**

2 *A. niger* assemblies, annotations, and analyzes are available through the interactive JGI
3 Genome Portal at <http://genome.jgi-psf.org/Aspni5/Aspni51.home.html>. Genome
4 assemblies together with predicted gene models and annotations were also deposited at
5 NCBI under the project accession number ACJE000000000.

6 **Strains**

7 The following *A. niger* strains were used for experiments (deposition numbers in different
8 collections are given as well): CBS 513.88 = FGSC A1513 = IBT 29270, ATCC 1015 = IBT 28639
9 = NCTC 3858a = NRRL 1278 = NRRL 350 = NRRL 511 = NRRL 328 = Thom 167 = CBS 113.46 =
10 ATCC 10582 = IMI 031821 = LSBH Ac4 = Thom 3528.7, NRRL 3 = IBT 28539 = MUCL 30480 =
11 DSM 2466 = CECT 2088 = VTT D-85240 = NRRL 566 = WB3 = ATCC 9029 = N400 = CBS 120.49
12 = IMI 041876, NRRL 326 = IBT 27876 = WB 326 = CBS 554.65 = ATCC 16888 = IHEM 3415 =
13 IMI 050566 = Thom 2766 = JCM 10254 = IFO 33023 (ex tannin-gallic acid fermentation,
14 Connecticut, USA) (*A. niger* neotype (Kozakiewicz et al. 1992)) NRRL 328 = IBT 27878 = NRRL
15 350 = IBT 27877 = CBS 113.46, NRRL 337 = CBS 126.48 = ATCC 10254 = DSM 734 = IFO 6428
16 = IMI 015954 = WB 337, NRRL 363 = IBT 3617 = IBT 5764 = CBS 126.49 = ATCC 10698 = IFO
17 6648, NRRL 511 = IBT 27875, NRRL 1278 = IBT 27872, NRRL 2270 = IBT 26391 = ATCC 11414
18 = VTT D-77050 = IMI 075353 = A60 = S.M. martin A-1-233 = Wisconsin 72-4, derived from
19 ATCC 1015, and NRRL 3122 = IBT 23538 = ATCC 22343 = CBS 115989 (This strain is a wild
20 type progenitor of CBS 513.88). Strain histories for ATCC 1015, ATCC 9029 and NRRL 3122
21 are reviewed by Baker (2006).

1 **Exo-metabolite profiling**

2 *A. niger* strains were inoculated on Czapek yeast autolysate (CYA), yeast extract sucrose
3 (YES) agar and CYA with 5% NaCl agar (CYAS agar). For medium composition see Frisvad and
4 Samson (2004). All strains were three-point inoculated on these media and incubated at
5 25°C in darkness for a week, after which 5 plugs (6 mm diameter) along a diameter of a
6 fungal colony were cut out and extracted (Smedsgaard 1997). The extracts were analyzed by
7 HPLC-DAD-flourescence (Frisvad and Thrane 1987; Smedsgaard 1997) and by HPLC-HR-MS
8 (Nielsen and Smedsgaard 2003; Frisvad et al. 2007b). The secondary metabolites were
9 identified by comparison with authentic standards (fumonisin B₂, ochratoxin A, nigragillin,
10 kotanin and orlandin) and by UV spectra and high-resolution mass spectra using
11 electrospray ionization.

12 **CBS 513.88 gap closure**

13 The near complete ATCC 1015 genome sequence was used to verify contig order and
14 orientation and to estimate gap sequence length between contigs in CBS 513.88. Gap
15 flanking PCR primers (20-mers) giving rise to PCR products with a minimal overlap of 100 bp
16 were automatically designed with primer-3 (Rozen and Skaletsky 2000) using custom scripts.
17 PCR products of expected size were purified using a QIAquick PCR Purification Kit (Qiagen
18 Inc, USA) and sequenced by Baseclear (Leiden, The Netherlands).

19 **Sequencing of variable regions**

20 Regions were amplified using standard PCR techniques with the four following primer pairs
21 (The name describes chromosome number and direction): 1) Chr02-Fwd: 5'-
22 GGAACTGCTTGATGTGATG-3', Chr02-Rev: 5'-GAGAGACGTACGAAAGGTTG-3'. 2) Chr04-Fwd:

1 5'-CGATCTGCGACCAGGA-3', Chr04-Rev: 5'-CATAACGGATTCGTCGCTG-3'. 3) Chr06-Fwd: 5'-
2 CTTGAAGGCGTTGAGGTC-3', Chr06-Rev: 5'-GCGAGTATGTGGCTAACATC-3'. 4) Chr08-Fwd: 5'-
3 GGTATGTACATTCCRTCCA-3', Chr08-Rev: 5'-GCTTGCAGTGAGCAAGGA-3'. The sequences on
4 Chrs II, VI, and VIII overlap with predicted genes. The PCR products were purified using a
5 QIAquick PCR Purification Kit (Qiagen Inc, USA) and sequenced by Agencourt Bioscience
6 Corporation (Beverly, Massachusetts). GenBank accession number GU296708-GU296739.

7 **Sequencing of β -tubulin**

8 Amplification of part of the β -tubulin gene was performed using the primers Bt2a
9 (GGTAACCAAATCGGTGCTGCTTTC) and Bt2b (ACCCTCAGTGTAGTGACCCTTGGC (Glass and
10 Donaldson 1995)). Both strands of the PCR fragments were sequenced with the ABI Prism®
11 Big Dye™ Terminator v.3.0 Ready Reaction Cycle sequencing Kit. Samples were analyzed
12 on an ABI PRISM 3700 Genetic Analyzer and contigs were assembled using the forward and
13 reverse sequences with the program SeqMan from the LaserGene package. GenBank
14 accession numbers GU296686-GU296707.

15 **Calculation of phylogenetic tree**

16 Sequences were aligned using ClustalX and trimmed to first common base at both ends
17 before making the final alignment. For the tree of Figure 3B, four trimmed sequences were
18 joined for each strain before making the final alignment. Tree calculations were performed
19 using TreeCon (van de Peer and de Wachter 1994). Distances were estimated using the
20 Jukes and Cantor algorithm taking insertions and deletions into account. Topology was
21 inferred using neighbor joining with bootstrap-values based on 1000 reiterations.

1 **Comparison of protein encoding loci**

2 The loci were compared using megablast of CBS 513.88 cDNA sequences Version 2008 with
3 standard setting followed by parsing of the output for local pair-wise sumscore of HSPs.

4 **Fermentation procedure**

5 *Growth media*

6 *A. niger* batch cultivation medium: 20 g/L glucose monohydrate, 7.3 g/L (NH₄)₂SO₄, 1.5 g/L
7 KH₂PO₄, 1.0 g/L MgSO₄·7H₂O, 1 g/L NaCl, 0.1 g/L CaCl₂·2H₂O, 0.05 mL/L antifoam 204
8 (Sigma) and 1 mL/L trace element solution. Trace element solution composition: 7.2 g/L
9 ZnSO₄·7H₂O, 0.3 g/L NiCl₂·6H₂O, 6.9 g/L FeSO₄·7H₂O, 3.5 g/L MnCl₂·4H₂O and 1.3 g/L
10 CuSO₄·5H₂O. *A. niger* complex medium: 2 g/L yeast extract, 3 g/L tryptone, 10 g/L glucose
11 monohydrate, 20 g/L agar, 0.52 g/L KCl, 0.52 g/L MgSO₄·7H₂O, 1.52 g/L KH₂PO₄ and 1 mL/L
12 of trace elements solution. Trace elements solution: 0.4 g/L CuSO₄·5H₂O, 0.04 g/L
13 Na₂B₄O₇·10H₂O, 0.8 g/L FeSO₄·7H₂O, 0.8 g/L MnSO₄·H₂O, 0.8 g/L Na₂MoO₄·2H₂O, 8 g/L
14 ZnSO₄·7H₂O.

15 *Spore propagation*

16 The bio-reactors were inoculated with spores of *A. niger* CBS 513.88 or ATCC 1015 strains
17 previously propagated on complex media plates and incubated for 8 days at 30°C. The same
18 stock of spores was used to inoculate all the plates. Spores were harvested by adding 10 mL
19 of Tween 80 0.01%, centrifuged at 4000 rpm for 10 min and washed three times with a
20 suitable amount of 0.9% NaCl. The fermentors were inoculated with a spore suspension to
21 obtain a final concentration of 5.7x10⁹ spores L⁻¹.

22 *Batch cultivations*

1 *A. niger* batch cultivations were carried out in 5 L reactors with a working volume of 4.5 L.
2 The bioreactors were equipped with two Rushton four-blade disc turbines, and pH and
3 temperature control. Inlet air was controlled with a mass flowmeter. The concentrations of
4 oxygen and carbon dioxide in the exhaust gas were monitored with a gas analyzer (1311
5 Fast response Triple gas, Innova combined with multiplexer controller for Gas Analysis
6 MUX100, B. Braun Biotech International). The temperature was maintained at 30°C and the
7 pH was controlled by automatic addition of 2 M NaOH. The pH was initially set to 3.0 to
8 prevent spore aggregation; when spores started to germinate, the pH was gradually
9 increased to 4.5. Similarly, the stirring speed was initially set to 200 rpm and the aeration
10 rate to 0.05 vvm (volume of gas per volume of liquid per minute) to prevent loss of
11 hydrophobic spores from the medium to the head-space of the reactor. After germination,
12 these parameters were progressively increased to 600 rpm and 0.89 vvm and kept steady
13 throughout all the rest of the fermentation.

14 *Sampling*

15 For quantification of cell mass and extracellular metabolites, the fermentation broth was
16 withdrawn from the reactor, filtered and washed. The filter cakes were used for cell weight
17 determination. The filtrate was filtered once more and frozen at -20°C for later HPLC
18 analysis. For gene expression analysis, mycelium was harvested in the exponential phase of
19 growth at half the maximum biomass density by filtration through sterile Miracloth
20 (Calbiochem, San Diego, CA, USA) and washed with a suitable amount of 0.9% NaCl solution.
21 The mycelium was quickly dried by squeezing and subsequently frozen in liquid nitrogen.
22 Samples were stored at -80°C until used for RNA extraction.

23 *Cell mass determination*

1 Cell dry weight was determined using paper filters (Whatmann cat. no. 1001070). The filters
2 were pre-dried in an oven at 100°C for 24 h, cooled down in a desiccator for 2 h and
3 subsequently weighed. A known volume of cell culture was filtered and the residue was
4 washed with 0.9% NaCl and dried on the filter for 24 h in an oven at 100°C. The filter was
5 weighed again and the cell mass concentration was calculated. Dry biomass was converted
6 to Cmol using 24.9 g biomass/Cmol (Nielsen et al. 2003).

7 *Quantification of sugars and extracellular metabolites*

8 The concentrations of sugar and organic acids in the filtrates were determined using HPLC
9 on an Aminex HPX-87H ion-exclusion column (BioRad, Hercules, CA). The column was eluted
10 at 60°C with 5 mM H₂SO₄ at a flow rate of 0.6 mL/min. Metabolites were detected with a
11 refractive index detector and an UV detector.

12 *Determination of glucoamylase activity*

13 The activity of glucoamylase in culture filtrates was quantified spectrophotometrically using
14 *p*-nitrophenyl β-maltoside as the substrate for glucoamylase (McCleary et al. 1991). 90 μl of
15 sodium acetate 200 mM (pH 4.4) were added to 10 μl of enzyme solution (standards or
16 fermentation broth), mixed and shaken for 30 s in a micro titer plate, then incubated 5 min
17 at 40°C. 10 μl of this mixture were mixed with 10 μl of substrate containing *p*-nitrophenyl β-
18 maltoside plus thermostable β-glucosidase (Megazyme). This preparation was shaken for 30
19 s and incubated for 10 min at 40°C. The reaction was stopped by the addition of 150 μl of
20 trizma base solution 2%. The sample preparation was shaken again for 30 seconds and
21 measured at 405 nm in a spectrophotometer. The absorbance measurement was related to
22 enzyme units per mL according to the calibration curve. One unit of enzyme is the amount
23 of enzyme which liberates 1 μmol of glucose per minute at pH 4.8 and 60°C. 1 unit can be

1 assumed to correspond to 25 µg protein (PESL protein assay; Boehringer Mannheim,
2 Mannheim, Germany).

3 **Transcriptome analysis**

4 *Extraction of total RNA, preparation of biotin-labeled cRNA and microarray processing*

5 Total RNA was isolated from 70–90 mg of frozen mycelium. 20 µg of fragmented biotin-
6 labeled cRNA was prepared from approximately 1 µg of total RNA. 15 µg of fragmented
7 cRNA was hybridized to the 3AspergDTU GeneChip, washed, stained, and scanned. The
8 scanned probe array images (.DAT files) were converted into .CEL files using the Affymetrix
9 GeneChip Operating Software. All steps were performed as in Andersen et al. (2008b)

10 *Analysis of transcription data*

11 Statistical analysis of the Affymetrix CEL-files was performed as described in Andersen et al.
12 (2008b). Note that this method uses the medianpolish algorithm (Irizarry et al. 2003), which
13 should make calculations of gene expression indexes relatively robust to differences in the
14 genome sequences. A cut-off value of adjusted $p < 0.05$ was set to assess statistical
15 significance. Normalized and raw data-values are deposited with GEO as series GSE10983.

16 **GO-term enrichment analysis**

17 *A. niger* ATCC 1015 ORF lists were examined for GO-term enrichment using R-2.5.1 (www.R-
18 project.org) with BioConductor (Gentleman et al. 2004) and the topGO-package v. 1.2.1
19 using the elim algorithm to remove local dependencies between GO terms (Alexa et al.
20 2006). GO-term assignments were based on automatic annotation of the *A. niger* ATCC 1015
21 gene models. Where nothing else is noted, $p < 0.05$ is used as the cutoff for significance.

1 **Genomic DNA hybridization analysis**

2

3 The data may be found in Suppl. Table 16. The dsmM_ANIGERa_coll511030F library and
4 platform information is deposited at GEO under number GPL6758.

5 **Genome-scale modeling**

6 Flux calculation was performed as described in Andersen et al. (2008a) using the *A. niger*
7 iMA871 model. Specifically, fluxes were calculated for ATCC 1015 and CBS 513.88 using the
8 growth conditions of the experiments described in Table 3, and using the concentrations of
9 glucose, glycerol, glucoamylase and citric acid from Table 3. The calculated fluxes may be
10 found in Suppl. Table 14.

11 **Synteny analysis**

12 The finished version of the *A. niger* ATCC 1015 genome sequence was divided into
13 fragments of 1 kb and compared to the nucleotide sequence of *A. niger* CBS 513.88 using
14 blastn (McGinnis and Madden 2004). A cutoff-value of 1e-75 was used. The blast results
15 were parsed using a custom made Perl script, giving for each 1 kb fragment the location of
16 the hit, number of SNPs, insertions and deletions. The same analysis was performed for a
17 comparison of the *A. niger* ATCC 1015 and *A. niger* ATCC 9029 genome sequences.

18 **Transcription mapping**

19 The nucleotide sequence of the gene models of the *A. niger* ATCC 1015 genome sequence
20 v1.0 were mapped to the 24 contigs of the finished sequence by blastn. A cutoff value of e-
21 100 was employed. An adapted version of the Imprint algorithm (see below) was employed

1 to find the coordinates of each gene. Log2-ratios of the average expression ratios in each of
2 the two strains were mapped to the coordinate of the first base of the gene finding a hit in
3 the target genome.

4 **Imprint algorithm**

5 The object of the algorithm is, for every coding sequence of *A. niger* ATCC 1015, to present
6 the corresponding genomic sequence from *A. niger* CBS 513.88 (the Imprint), so that an un-
7 biased comparison of two sets of gene callers can be conducted.

8 The CDSs of *A. niger* ATCC 1015 were compared to the nucleotide sequence of *A. niger* CBS
9 513.88 using blastn (McGinnis and Madden 2004). The blastn output was parsed to extract -
10 for each gene - the alignment to the *A. niger* CBS 513.88 sequence. This provided for each
11 gene, the part of the CDS having a hit in the CBS 513.88 sequence, and the corresponding
12 sequence from CBS 513.88 sequence (the Imprint). The CDS and the Imprint were compared
13 to produce a comprehensive list of insertions, deletions, silent mutations, sense mutations,
14 frame shift mutations and non-sense mutations for each CDS-pair. Detail on the analysis and
15 manual curation of the results can be found in Suppl. Text 3. The final summary is found in
16 Suppl. Table 5. All genes were assigned to one of the categories of Suppl. Table 5. Genes
17 were only assigned to categories E–F if none of the requirements of categories G–L were
18 fulfilled. Manual and automatic annotation was added for 9,400 genes, including all genes
19 with more than 0.5% difference in amino acid sequence.

20 ***Acknowledgements***

21 This study was partially funded by the Danish Research Agency for Technology and
22 Production (MRA, SLM, JBN, MLN) and the National Council of Research Conacyt-Mexico

1 (MPS). Genome sequencing was performed under the auspices of the US Department of
2 Energy's Office of Science, Biological and Environmental Research Program, and by the
3 University of California, Lawrence Berkeley National Laboratory under contract No. DE-
4 AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-
5 07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.
6 Preparation of genomic DNA and some analysis of the genome were supported by the DOE
7 Office of the Biomass Program.

8 ***Figure legends***

9 **Figure 1**

10 Synteny map of the contigs of *A. niger* ATCC 1015 to the supercontigs of *A. niger* CBS
11 513.88. The coloring of the chromosomes shows syntenic regions in *A. niger* CBS 513.88.
12 Arabic numerals show the number of the supercontig in *A. niger* CBS 513.88. Gray areas
13 shows regions not found in the CBS 513.88 genome sequence (Pel et al. 2007). Proposed
14 locations of centromeric regions are shown with filled black circles. Sequenced telomeres
15 are marked with a T. Zeroes mark the first base of the contigs. A black line underneath a
16 section of the chromosomes denotes inverted sequence. Black histograms show SNPs/kb
17 (number of single nucleotide polymorphisms/kb) between the sequences of the two strains
18 (Y-axis: 0–160 SNPs/kb). Gaps between contigs and centromeres are not to scale. The
19 alignment demonstrates almost complete synteny between the two strains, with the
20 exception of a cross-over event between the left arms of Chrs III and VII. An overview of the
21 genes found in the gaps unique to ATCC 1015 can be found in Suppl. Table 2.

1 **Figure 2**

2 Horizontal gene transfer of α -amylase genes from *A. oryzae* to *A. niger* CBS 513.88.

3 **(A)** Unmatched region identified for the left arm of Chr III spanning 65 kb for ATCC 1015
4 encoding 30 predicted genes and 85 kb for CBS 513.88 encoding 24 predicted genes.
5 The unmatched region is flanked on one side by a small local inversion of 3 predicted genes
6 (in green).

7 **(B)** The 12.4 kb HGT region is part of an identical 12.7 kb region present in *A. oryzae* RIB40
8 supercontig SC113 and SC023. In *A. niger* CBS 513.88, the transferred region is enclosed by a
9 203 bp inverted repeat (red arrow). An12g07000 (blue) is a putative transposase and
10 identical to the *A. oryzae* transposon Aot1 tnpA gene. Similarly the *A. niger* CBS 513.88 α -
11 amylase encoding gene An12g06930 (orange) is identical to *A. oryzae* annotated genes
12 AO090120000196 (SC113) and AO090023000944 (SC023).

13 **(C)** Proposed duplication - recombination event between supercontig 12 and supercontig 05
14 of the 12 kb HGT region. Breakpoints are indicated with dotted lines. The fragment encoding
15 α -amylase An05g02200 is identical to An12g06930 (orange) The breakpoints are flanked
16 with additional copies of the 203 bp repeat region (red arrow). The region encoding genes
17 An12g06940 to An12g06970 is identical to the region encoding An05g02210 to An05g02130
18 (green arrows). The downstream breakpoint occurred in the predicted gene coding region of
19 An12g06970, thereby deleting the original start codon and upstream region.

20 **Figure 3**

21 **(A)** Phylogenetic relationship of several black *Aspergilli* based on partial sequencing of β -
22 tubulin. The tree was rooted to *A. aculeatus* CBS 172.66.

1 **(B)** Phylogenetic relationship of seven strains of *A. niger* based on sequencing of 1 kb
2 variable regions from four chromosomes. The tree was rooted to the sequence obtained
3 from *A. carbonarius* IMI 388653. Clades based on the exo-metabolomic groupings of Table 2
4 are shown.

5 For both trees, bootstrap values above 80 % of the 1000 performed re-iterations are shown.

6

1 *Tables*

2 **Table 1:** General genome statistics for *A. niger* ATCC 1015 and *A. niger* CBS 513.88. Except
3 genome sizes and the number of gene models, all values are averages.

	<i>A. niger</i> ATCC 1015	<i>A. niger</i> CBS 513.88^a	<i>A. niger</i> CBS 513.88^b
	(this study)	(Pel et al. 2007)	(this study)
Gene models	11,200	14,165	14,082
Genome size (Mbp)	34.85	33.93	34.02
Protein length (aa)	484.3	439.9	442.5
Exons per gene	3.1	3.6	3.6
Exon length (bp)	480.8	370.0	371.6
Intron length (bp)	93.8	97.2	96.9

4 ^aThe genome assembly published by Pel et al. (2007).

5 ^bGenome assembly of *A. niger* CBS 513.88 after gap closure using sequence information
6 from ATCC 1015.

7

1 **Table 2:** Exo-metabolomic profiling of 11 *A. niger* strains based on HPLC-DAD-FLD and HPLC-
2 DAD-HRMS (this study). The reference column relates to the elucidation of the compound
3 structure.

Secondary metabolites	Reference	NRRL 3122	CBS 513.88 ^a	CBS 126.49	ATCC 1015 ^a	ATCC 9029 ^a	CBS 554.65	NRRL 328	NRRL 350	NRRL 511	NRRL 1278	NRRL 2270
Aurasperone B	(Tanaka et al. 1966)	•	•	•	•	•	•	•	•	•	•	•
Fumonisin B ₂	(Frisvad et al. 2007b)	•	•		•	•	•	•	•	•	•	•
Funalenone	(Inokoshi et al. 1999)			•	•	•	•	•	•	•	•	•
Kotanin	(Büchi et al. 1971)			•	•	•	•	•	•	•	•	•
Nigragillin	(Isogai et al. 1975)	•	•									
Ochratoxin A	(Abarca et al. 1994)	•	•	•								
Orlandin	(Cutler et al. 1979)			•	•	•	•	•	•	•	•	•
Other naphtho-γ-pyrones				•	•	•	•	•	•	•	•	•
Pyranonigrin A	(Hiort et al. 2004)	•	•	•	•	•	•	•	•	•	•	•
Tensidol B	(Fukada et al. 2006)	•	•	•	•	•	•	•	•	•	•	•

4 ^aGenome-sequenced strains

5

Table 3: Statistics for batch cultivations of *A. niger* ATCC 1015 and *A. niger* CBS 513.88. Fermentations were performed in biological triplicates for each strain. Values are presented as average±standard deviation. μ_{\max} and Y_{xs} are general statistics for the fermentations, while the remaining values are specific for the time of sampling for transcription analysis (see mRNA row). GlA is glucoamylase A.

	ATCC 1015	CBS 513.88
mRNA (h)	24.5±1.2	40.2±4.2
Biomass (g/L)	5.0±0.1	4.0±0.5
μ_{\max} (h ⁻¹)	0.17±0.01	0.15±0.01
Glucose (g/L)	10.0±0.6	9.5±0.4
Glycerol (g/L)	0.09±0.02	0.27±0.03
Y_{xs} (Cmol/Cmol) ^a	0.67±0.03	0.55±0.03
GlA (g/L) ^b	0.24±0.08	1.57±0.23
Citric acid (g/L)	0.10±0.12	0.14±0.03

6

^aBiomass was converted to Cmol using 24.9 g biomass/Cmol (Nielsen et al. 2003).

^bOne unit of glucoamylase can be assumed to correspond to 25µg of protein (PESL protein assay; Boehringer Mannheim, Mannheim, Germany).

10

1 **References**

- 2 Abarca ML, Bragulat MR, Castella G, Cabanes FJ. 1994. Ochratoxin A production by strains of
3 *Aspergillus niger* var. *niger*. *Appl. Environ. Microbiol.* **60**: 2650–2652.
- 4 Alexa A, Rahnenführer J, Lengauer T. 2006. Improved scoring of functional groups from gene
5 expression data by decorrelating GO graph structure. *Bioinformatics* **22**: 1600–1607.
- 6 Amutan M, Nyyssonen E, Stubbs J, Diaz-Torres MR, Dunn-Coleman N. 1996. Identification
7 and cloning of a mobile transposon from *Aspergillus niger* var. *awamori*. *Curr. Genet.* **29**:
8 468–473.
- 9 Andersen MR, Nielsen M, Nielsen J. 2008a. Metabolic model integration of the bibliome,
10 genome, metabolome and reactome of *Aspergillus niger*. *Mol. Syst. Biol.* **4**: 178. doi:
11 10.1038/msb.2008.12.
- 12 Andersen MR, Vongsangnak W, Panagiotou G, Salazar M, Lehmann L, Nielsen J. 2008b. A tri-
13 species *Aspergillus* micro array: comparative transcriptomics of three *Aspergillus* species.
14 *Proc. Natl. Acad. Sci. U S A* **105**: 4387–4392.
- 15 Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S,
16 Smit A, et al. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu*
17 *rubripes*. *Science* **297**: 1301–1310.
- 18 Baker SE. 2006. *Aspergillus niger* genomics: past, present and into the future. *Med. Mycol.*
19 **44 Suppl 1**: S17-21.
- 20 Bos CJ, Debets AJ, Kobus G, Slakhorst SM, Swart K. 1989. Adenine and pyrimidine genes of
21 *Aspergillus niger* and evidence for a seventh linkage group. *Curr. Genet.* **16**: 307–310.

1 Braumann I, van den Berg M, Kempken F. 2007. Transposons in biotechnologically relevant
2 strains of *Aspergillus niger* and *Penicillium chrysogenum*. *Fungal. Genet. Biol.* **44**: 1399-1414.

3 Braumann I, van den Berg M, Kempken F. 2008. Strain-specific retrotransposon-mediated
4 recombination in commercially used *Aspergillus niger* strain. *Mol. Genet. Genomics* **280**:
5 319-325.

6 Büchi G, Klaubert D, Shank R, Weinreb S, Wogan GN. 1971. Structure and synthesis of
7 kotanin and desmethylokotannin, metabolites of *Aspergillus glaucus*. *J. Org. Chem.* **36**: 1143–
8 1147.

9 Christias C, Couvaraki C, Georgopoulos SG, Vomvoyanni V. 1975. Protein content and amino
10 acid composition of certain fungi evaluated for microbial protein production. *Appl.*
11 *Microbiol.* **29**: 250–254.

12 Cullen D. 2007. The genome of an industrial workhorse. *Nat. Biotechnol.* **25**: 189–190.

13 Cuomo CA, Güldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma LJ, Baker SE,
14 Rep M, et al. 2007. The *Fusarium graminearum* genome reveals a link between localized
15 polymorphism and pathogen specialization. *Science* **317**: 1400–1402.

16 Cutler HG, Crumley FG, Cox RH, Hernandez O, Cole RJ, Dorner JW. 1979. Orlandin: a
17 nontoxic fungal metabolite with plant growth inhibiting properties. *J. Agric. Food Chem.* **27**:
18 592–595.

19 Debets A, Holub E, Swart K, van den Broek H, Bos C. 1990a. An electrophoretic karyotype of
20 *Aspergillus niger*. *Mol. Gen. Genet.* **224**: 264–268.

21 Debets A, Swart K, Bos C. 1989. Mitotic mapping in linkage group V of *Aspergillus niger*
22 based on selection of auxotrophic recombinants by Novozym enrichment. *Can. J. Microbiol.*
23 **35**: 982–988.

1 Debets A, Swart K, Bos C. 1990b. Genetic analysis of *Aspergillus niger*: isolation of chlorate
2 resistance mutants, their use in mitotic mapping and evidence for an eighth linkage group.
3 *Mol. Gen. Genet.* **221**: 453–458.

4 Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, Crabtree J, Silva JC,
5 Badger JH, Albarraq A, et al. 2008. Genomic islands in the pathogenic filamentous fungus
6 *Aspergillus fumigatus*. *PLoS Genet.* **4**: e1000046. doi: 10.1371/journal.pgen.1000046.

7 Frisvad JC, Thrane U. 1987. Standardized High Performance Liquid Chromatography of 182
8 mycotoxins and other fungal metabolites based on alkylphenone indices and UV VIS spectra
9 (diode array detection). *J. Chromatogr.* **404**: 195-214.

10 Frisvad JC, Samson R. 2004. Polyphasic taxonomy of *Penicillium* subgenus *Penicillium*. A
11 guide to identification of the food and air-borne terverticillate *Penicillia* and their
12 mycotoxins. *Stud. Mycol.* **49**: 1–173.

13 Frisvad JC, Larsen TO, de Vries R, Meijer M, Houbraken J, Cabañes FJ, Ehrlich K, Samson RA.
14 2007a. Secondary metabolite profiling, growth profiles and other tools for species
15 recognition and important *Aspergillus* mycotoxins. *Stud. Mycol.* **59**: 31-37.

16 Frisvad JC, Smedsgaard J, Samson R, Larsen T, Thrane U. 2007b. Fumonisin B2 production by
17 *Aspergillus niger*. *J. Agric. Food Chem.* **55**: 9727–9732.

18 Fukuda T, Hasegawa Y, Hagimori K, Yamaguchi Y, Masuma R, Tomoda H, Omura S. 2006.
19 Tensidols, new potentiators of antifungal miconazole activity, produced by *Aspergillus niger*
20 FKI-2342. *J. Antibiot. (Tokyo)* **59**: 480–485.

21 Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, Lee SI, Baştürkmen M,
22 Spevak CC, Clutterbuck J, et al. 2005. Sequencing of *Aspergillus nidulans* and comparative
23 analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438**: 1105–1115.

1 Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y,
2 Gentry J, et al. 2004. Bioconductor: Open software development for computational biology
3 and bioinformatics. *Genome Biol.* **5**: R80. doi: 10.1186/gb-2004-5-10-r80.

4 Glass NL, Donaldson GC. 1995. Development of primer sets designed for use with the PCR to
5 amplify conserved genes from filamentous ascomycetes. *Appl. Environ. Microbiol.* **61**: 1323-
6 1330.

7 Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing.
8 *Genome Res.* **8**: 195-202.

9 Gottschling D, Aparicio O, Billington B, Zakian V. 1990. Position effect at *S. cerevisiae*
10 telomeres: reversible repression of Pol II transcription. *Cell* **63**: 751–762.

11 Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B,
12 Mungall C, et al. 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic*
13 *Acids Res.* **1**: D258-61.

14 Hiort J, Maksimenka K, Reichert M, Perović-Ottstadt S, Lin WH, Wray V, Steube K,
15 Schaumann K, Weber H, Proksch P, et al. 2004. New natural products from the sponge-
16 derived fungus *Aspergillus niger*. *J. Nat. Prod.* **67**: 1532–1543.

17 Inokoshi J, Shiomi K, Masuma R, Tanaka H, Yamada H, Omura S. 1999. Funalenone, a novel
18 collagenase inhibitor produced by *Aspergillus niger*. *J. Antibiot. (Tokyo)* **52**: 1095–1100.

19 Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. 2003.
20 Exploration, normalization, and summaries of high density oligonucleotide array probe level
21 data. *Biostatistics* **4**: 249–64.

- 1 Isogai A, Horii T, Suzuki A, Murakoshi S, Ikeda K, Sato S, Tamura S. 1975. Isolation and
2 identification of nigragillin as an insecticidal metabolite produced by *Aspergillus niger*. *Agric.*
3 *Biol. Chem.* **39**: 739–740.
- 4 Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for
5 deciphering the genome. *Nucleic Acids Res* **32**: D277–280.
- 6 Kanehisa M, Guto S, Kawashima S, Nakaya A. 2002. The KEGG databases at GenomeNet.
7 *Nucleic Acids Res.* **30**: 42–46.
- 8 Karaffa L, Kubicek CP. 2003. *Aspergillus niger* citric acid accumulation: do we understand this
9 well working black box? *Appl. Microbiol. Biotechnol.* **61**: 189–196.
- 10 Keeling PJ, Palmer, JD. 2008. Horizontal gene transfer in eukaryotic evolution. *Nat. Rev.*
11 *Genet.* **9**: 605-617.
- 12 Khaldi N, Collemare J, Lebrun MH, Wolfe KH. 2008. Evidence for horizontal transfer of a
13 secondary metabolite gene cluster between fungi. *Genome Biol.* **9**: R18. doi: 10.1186/gb-
14 2008-9-1-r18.
- 15 Khaldi N, Wolfe KH. 2008. Elusive origins of the extra genes in *Aspergillus oryzae*. *PLoS ONE*,
16 **3**: e3036. doi: 10.1371/journal.pone.0003036.
- 17 Korman DR, Bayliss FT, Barnett CC, Carmona CL, Kodama KH, Royer TJ, Thompson SA, Ward
18 M, Wilson LJ, Berka RM. 1990. Cloning, characterization, and expression of two alpha-
19 amylase genes from *Aspergillus niger* var. *awamori*. *Curr. Genet.* **17**: 203-12.
- 20 Kozakiewicz Z, Frisvad JC, Hawksworth DL, Pitt JI, Samson RA, Stolk AC. Proposals for nomina
21 specifica conservanda and rijicienda in *Aspergillus* and *Penicillium* (Fungi). *Taxon* **41**: 109–
22 113.

- 1 Lubertozzi D, Keasling JD. 2008. Developing *Aspergillus* as a host for heterologous
2 expression. *Biotechnol. Adv.* **27**: 53-75.
- 3 Martens-Uzunova ES, Schaap PJ. 2008. An evolutionary conserved D-galacturonic acid
4 metabolic pathway operates across filamentous fungi capable of pectin degradation. *Fungal*
5 *Genet. Biol.* **45**: 1449-1457.
- 6 McCleary BV, Bouhet F, Driguez, H. 1991. Measurement of amyloglucosidase using p-
7 nitrophenyl b-maltoside as substrate. *Biotech. Tech.* **5**: 255-258.
- 8 McGinnis S, Madden T. 2004. BLAST: at the core of a powerful and diverse set of sequence
9 analysis tools. *Nucleic Acids Res.* **32**: W20–25.
- 10 Nielsen J, Villadsen J, Liden G. 2003. *Bioreaction Engineering Principles*, 2nd edition. Kluwer
11 Academic/Plenum Publishers, New York.
- 12 Nielsen KF, Smedsgaard J. 2003. Fungal metabolite screening: database of 474 mycotoxins
13 and fungal metabolites for dereplication by standardized liquid chromatography-UV-mass
14 spectrometry methodology. *J. Chromatogr. A.* **1002**: 111–136.
- 15 Pel HJ, de Winde JH, Archer DB, Dyer PS, Hofmann G, Schaap PJ, Turner G, de Vries RP,
16 Albang R, Albermann K, et al. 2007. Genome sequencing and analysis of the versatile cell
17 factory *Aspergillus niger* CBS 513.88. *Nat. Biotechnol.* **25**: 221–231.
- 18 Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist
19 programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds.
20 S Krawetz and S Misener), pp. 365-386. Humana Press, Totowa, NJ.
- 21 Schuster E, Dunn-Coleman N, Frisvad JC, van Dijck PWM. 2002. On the safety of *Aspergillus*
22 *niger* – a review. *Appl. Microbiol. Biotechnol.* **59**: 426-435.

- 1 Shibuya I, Tamura G, Ishikawa T, Hara S. 1992. Cloning the alpha-amylase cDNA of
2 *Aspergillus shirousamii* and its expression in *Saccharomyces cerevisiae*. *BioSci. Biotech.*
3 *Biochem.* **56**: 174-179.
- 4 Smedsgaard J. 1997. Micro-scale extraction procedure for standardised screening of fungal
5 metabolite production in cultures. *J. Chromatogr. A.* **760**: 264–270.
- 6 Strand DJ, McDonald JF. 1985. Copia is transcriptionally responsive to environmental stress.
7 *Nucleic Acids Res.* **13**:4401-10.
- 8 Sun J, Lu X, Rinas U, Zeng A. 2007. Metabolic peculiarities of *Aspergillus niger* disclosed by
9 comparative metabolic genomics. *Genome Biol.* **8**: R182. doi: 10.1186/gb-2007-8-9-r182.
- 10 Swart K, Debets AJ, Kobus G, Bos CJ. 1992. Arginine and proline genes of *Aspergillus niger*.
11 *Antonie Van Leeuwenhoek* **61**: 259–264.
- 12 Tanaka H, Wang PL, Yamada O, Tamura L. 1966. Yellow pigments of *Aspergillus niger* and
13 *Asp. awamori*. Part. I. Isolation of aurasperone A and related pigments. *Agric. Biol. Chem.*
14 **30**: 107–113.
- 15 van de Peer Y, de Wachter R. 1994. TREECON for Windows: a software package for the
16 construction and drawing of evolutionary trees for the Microsoft Windows environment.
17 *Comput. Appl. Biosci.* **10**: 569-70.
- 18 van Dijck PWM. 2008. The importance of *Aspergilli* and regulatory aspects of *Aspergillus*
19 nomenclature. In: *Aspergillus in the genomic era* (eds. J Varga and RA Samso), pp. 249-57.
20 Wageningen Academic Publishers, Wageningen, The Netherlands.
- 21 van Dijck PWM, Selten GCM, Hempenius RA. 2003. On the safety of a new generation of
22 DSM *Aspergillus niger* production strains. *Regul. Toxicol. Pharmacol.* **38**: 27-35.

- 1 van Lanen JM, Smith MB. 1968. Process of producing glucamylase and an alcohol product.
2 US patent 3,418,211. Hiram Walker & Sons, Inc., Peoria, IL.
- 3 Verdoes JC, Calil MR, Punt PJ, Debets F, Swart K, Stouthamer AH, van den Hondel CA. 1994.
4 The complete karyotype of *Aspergillus niger*: the use of introduced electrophoretic mobility
5 variation of chromosomes for gene assignment studies. *Mol. Gen. Genet.* **244**: 75–80.
- 6 Yuan X, Roubos J, van den Hondel C, Ram A. 2008a. Identification of InuR, a new Zn(II)2Cys6
7 transcriptional activator involved in the regulation of inulinolytic genes in *Aspergillus niger*.
8 *Mol. Genet. Genomics* **279**: 11–26.
- 9 Yuan XL, van der Kaaij RM, van den Hondel CA, Punt PJ, van der Maarel MJ, Dijkhuizen L,
10 Ram AF. 2008b. *Aspergillus niger* genome-wide analysis reveals a large number of novel α -
11 glucan acting enzymes with unexpected expression profiles. *Mol. Genet. Genomics* **279**:
12 545–561.